Chapter 156

Tests for the Odds Ratio in a Matched Case-Control Design with a Binary X

Introduction

This procedure calculates the power and sample size necessary in a matched case-control study designed to detect a relationship between the development of a disease and a risk factor (exposure variable) using an odds ratio computed from a conditional logistic regression. The procedure also provides an adjustment to power for other covariates. Kleinbaum and Klein (2010) provide a detailed discussion of the interpreting the odds ratio in a conditional logistic regression.

Suppose a subject population is to be studied for the relationship between an outcome variable (such as lung cancer) and a binary risk factor (such as cigarette smoking). A matched case-control study is planned in which N matched sets will used. Each matched set will consist of M_D *case* subjects which are positive for the outcome (diseased) and M_H *control* subjects that are negative for the outcome (healthy). The subjects in each set are matched according other covariates that are assumed to have a large impact on the probability of the disease such as age and gender. In each matched set, some subjects are positive for a binary *exposure* variable of interest. Note that the design may be *retrospective* or *prospective*.

Technical Details

Hypotheses are investigated using a score test of the log odds ratio in a conditional logistic regression. Power and sample size formulas are given in Lachin (2008) and Tang (2009).

Consider a binary exposure variable E which is set to '1' if the patient has been exposed to the risk factor of interest or to '0' if not. The probability that a case patient was exposed to the risk factor is P_{ED} and the probability that a control patient was exposed to the risk factor is P_{EH}. Furthermore, the probability that any patient has been exposed to the risk factor is P_E. Hence, P_E can be thought of as the population prevalence of exposure. If you don't have better information, you can use $P_E = (P_{ED} + P_{EH})/2$ or, if the disease is rare, $P_E = P_{EH}$.

We consider that a score test will be constructed from a conditional logistic regression to test hypotheses about the regression coefficient corresponding to the exposure variable which we will call θ (see Lachin (2008) for details). For reasonable sample sizes, this score test can be assumed to follow the standard normal distribution.

The regression coefficient θ of the exposure variable in a conditional logistic regression is interpreted as the log odds ratio. Note that in a case-control study, the odds ratio is defined as

$$OR = e^\theta = \frac{P_{ED}/(1 - P_{ED})}{P_{EH}/(1 - P_{EH})}$$

It important to note that the numerator in this expression gives the odds that case patient was exposed and the denominator gives the odds that a control patient was exposed. It is useful to recognize that these are the odds that a subject has been exposed, not the odds that the subject has the disease. Hence, OR is the ratio of odds that a case subject has been exposed to the odds that a control subject has been exposed.

The power of the score test is calculated using

$$z_{1-\beta} = |\theta| \sqrt{NP_E(1 - P_E)\left(\frac{M_D M_H}{M_D + M_H}\right)} - z_{1-\alpha}$$

This can be rearranged to obtain the following expression for sample size

$$N = \frac{\left(z_{1-\beta} + z_{1-\alpha}\right)^2}{\theta^2 P_E(1 - P_E)\left(\frac{M_D M_H}{M_D + M_H}\right)}$$

Note that for two-sided tests, α is replaced by α/2.

Adjusting for Other Covariates

Lachin (2008) provides an adjustment to the power when additional covariates are fit in the conditional logistic regression. Let $R^2_{X|Z}$ represent the coefficient of determination for a (multiple) regression of the exposure variable X on the covariates Z. Note that Z is a vector of 1 or more covariates and that the number of covariates is not needed. The adjustment is made by multiplying $P_E(1 - P_E)$ by $1 - R^2_{X|Z}$ in the formulas above. Lachin stresses that in order for this adjustment to be accurate, none of the adjusting covariates can have a strong effect upon the response. He indicates that this assumption should be meet since any covariate with a large effect should be controlled for by the matching.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. Select *Sample Size* when you want to calculate the *Sample Size* needed to achieve a given power and alpha level.

Select *Power* when you want to calculate the power of an experiment that has already been run.

Test

Alternative Hypothesis

Specify whether the alternative hypothesis of the test is one-sided or two-sided. If a one-sided test is chosen, the hypothesis test direction is chosen based on whether OR is greater than or less than one.

Two-Sided Hypothesis Test

H0: OR = 1 vs. Ha: OR $\neq$ 1

One-Sided Hypothesis Tests

Upper: H0: OR $\leq$ 1 vs. Ha OR > 1

Lower: H0: OR $\geq$ 1 vs. Ha OR < 1

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

Alpha is the probability of rejecting a true null hypothesis. The null hypothesis is that the odds ratio (OR) is 1. This means that the probability of a positive outcome is no different for those that have been exposed to the risk factor and those that have not.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

N (Number of Sets or Strata)

N is the number of matched sets in the study. Each set consists of M_D cases and M_H controls. The total number of subjects is N x (M_D + M_H).

The range of possible values is of 3 or more. Multiple values may be entered.

M_D (Number of Cases per Set)

The number of case subjects in each of the N matched sets. The cases are those subjects with a positive outcome, such as diseased. Thus, the total number of subjects is N x (M_D + M_H).

Values of one or greater are allowed. Seldom is this value chosen to be more than five.

Multiple values may be entered.

M (Number of Controls per Set)

The number of control subjects in each of the N matched sets. The controls are those subjects with a negative outcome, such as healthy (non-diseased). The total number of subjects is N x (M_D + M_H).

Values of one or greater are allowed. Seldom is this value chosen to be more than ten.

Multiple values may be entered.

Effect Size

OR (Odds Ratio)

The value of the odds ratio to be detected. This is the odds ratio at which the power is calculated. Note that the power or sample size will be identical for 1/OR and OR.

The odds ratio is the ratio of the odds of an exposed subject having the disease to the odds of an unexposed subject having the disease.

The odds ratio may also be interpreted as is the ratio of the odds of a diseased subject having been exposed to the risk factor to the odds of a healthy subject having been exposed.

For example, an odds ratio of 2.0 means that subjects who were exposed to the risk factor have twice the odds of having the disease as do unexposed subjects.

Range

A value greater than one is usually used. The value must be greater than zero. The null hypothesis is that the odds ratio is one. It is best to use a value that is large enough to be of interest to others such as 1.5 or 2.0.

P_E (Probability of Exposure)

This is the probability that a subject selected at random from the general population of interest has been exposed to the risk factor. You can approximate P_E with (P_{ED} + P_{EH})/2, of, if E is rare, with P_{EH}. Here, P_{ED} is the probability that a case subject is exposed and P_{EH} is the probability that a control subject is exposed.

The valid range is between 0 and 1.

You can enter a single value such as *0.05*, a list of values such as *0.1 0.3 0.5*, or a series of values such as

0.05 to 0.50 by 0.05.

R^2 (Exposure vs. Covariates)

Enter the R^2 value that would be obtained if the Exposure variable was regressed on the other covariates (independent variables) to be used in the conditional logistic model. This is referring to the R^2 from a regular multiple regression of the exposure variable on the covariates, not a logistic regression.

This R^2 value is used to calculate an adjustment to the power and sample size when the covariates are in the model. Note that the number of covariates is not used in the adjustment and so it is not entered.

If you do not have any covariates, enter '0.0'.

The absolute range is from zero to just less than one. However, it is assumed that covariates with large impacts will be used in the matching process.

You can enter a single value such as *0.1*, a list of values such as *0.05 0.1 0.2*, or a range of values such as *0.0 to 0.1 by 0.01*.

Example 1 – Calculating Sample Size

This example will show how to calculate the power of a two-sided, retrospective study for several sample sizes and odds ratios.

Suppose that a matched case-control study is to be run in which the OR = 1.5, 2.0, 2.5, or 3.0, $P_E = 0.3$, $R^2 = 0.2$, $M_D = 1$, $M_H = 1$, 2, or 5, power = 0.9, and alpha = 0.05, and power is to be found.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Odds Ratio in a Matched Case-Control Design with a Binary X** procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size N**
Alternative Hypothesis	**Two-Sided**
Power	**0.90**
Alpha	**0.05**
M_D (Number of Cases per Set)	**1**
M_H (Number of Controls per Case)	**1 2 5**
OR (Odds Ratio)	**1.5 2 2.5 3**
P_E (Probability of Exposure)	**0.3**
R^2 (Exposure vs. Covariates)	**0.2**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results
Hypothesis Type: Two-Sided

Power	Number of Matched Sets N	Cases per Set M_D	Controls per Set M_H	Odds Ratio OR	Probability of Risk Exposure P_E	Regression of Exposure on other Covariates R^2	Alpha
0.9000	761	1	1	1.5000	0.3000	0.2000	0.0500
0.9000	571	1	2	1.5000	0.3000	0.2000	0.0500
0.9000	457	1	5	1.5000	0.3000	0.2000	0.0500
0.9000	261	1	1	2.0000	0.3000	0.2000	0.0500
0.9000	196	1	2	2.0000	0.3000	0.2000	0.0500
0.9000	157	1	5	2.0000	0.3000	0.2000	0.0500
0.9000	149	1	1	2.5000	0.3000	0.2000	0.0500
0.9000	112	1	2	2.5000	0.3000	0.2000	0.0500
0.9000	90	1	5	2.5000	0.3000	0.2000	0.0500
0.9000	104	1	1	3.0000	0.3000	0.2000	0.0500
0.9000	78	1	2	3.0000	0.3000	0.2000	0.0500
0.9000	63	1	5	3.0000	0.3000	0.2000	0.0500

References
Lachin, John M. 2008. 'Sample size evaluaion for a multiply matched case-control study using the score test
    from a conditional logistic (discrete Cox PH) regression model.' Statistics in Medicine, Volume 27, Pages
    2509-2523.
Lachin, John M. 2011. Biostatistical Methods: The Assessment of Relative Risks, Second Edition. John Wiley &
    Sons. New York.
Tang, Yongqiang. 2009. 'Comments on 'Sample size evaluation for multiply matched case-control study using the
    score test from a conditional logistic (discrete Cox PH) regression model.'' Statistics in Medicine, Volume
    28, Pages 175-177.

Report Definitions
Power is the probability of rejecting a false null hypothesis.
Number of Matched Sets, N, is the number of sets (strata) in the study. Each set consists in a fixed number of
    cases and controls.
Cases per Set, M_D, is the number cases in each matched set.
Controls per Set, M_H, is the number of controls in each matched set.
Odds Ratio, OR, is the odds ratio of developing a disease associated with exposure to a certain risk factor.
Probability of Risk Exposure, P_E, is the probability of exposure to the risk factor in the overall population.
Regression of Exposure on other Covariates, R^2, is the R^2 that occurs when the exposure variable is regressed
    on any other covariates. This adjustment assumes that covariates that have a large correlation with the
    outcome are used in the matching process and are not included here.
Alpha is the probability of rejecting a true null hypothesis of no association between disease and the
    exposure variable.

Summary Statements
In a matched case-control study, a sample of 761 matched sets (or strata) is obtained. Each
matched set consists of 1 case and 1 control. The probability of exposure to the risk factor in
the population is 0.3000. This sample achieves 90% power to detect an odds ratio of 1.5000
calculated using conditional logistic regression with a 0.0500 significance level.

This report shows the power for each of the scenarios.

Plots Section
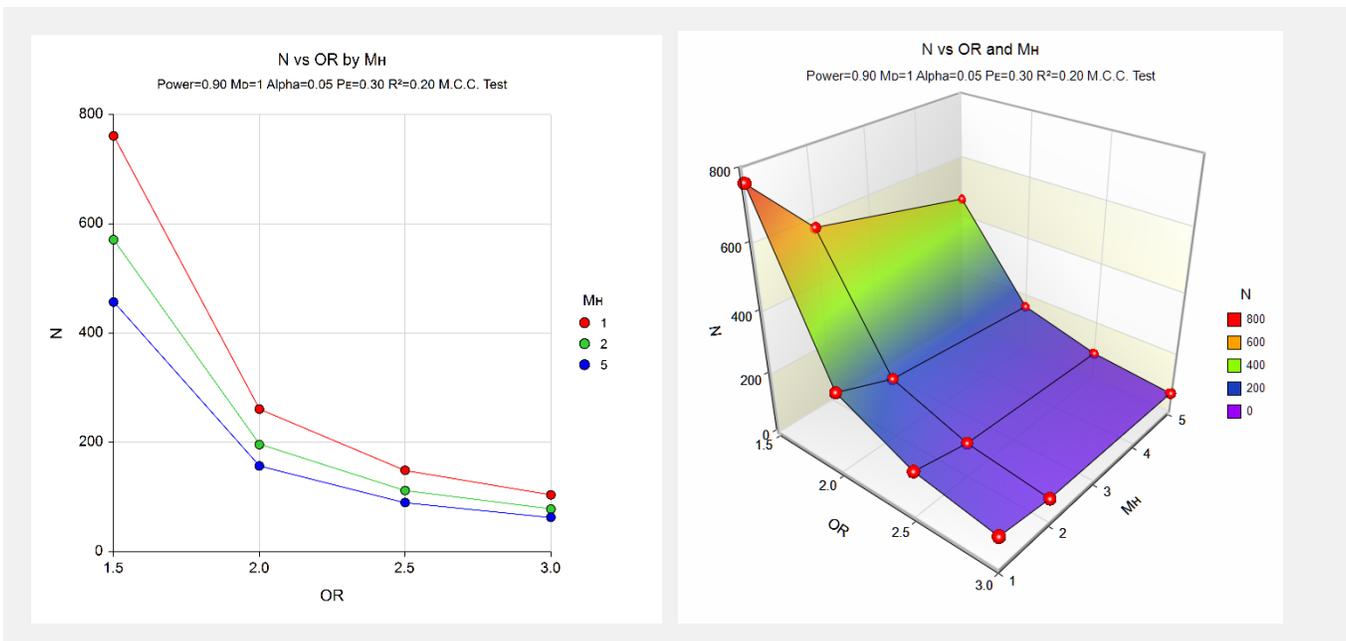
This plot shows the sample size versus the odds ratio for the three M_H's.

Example 2 – Validation using Lachin (2011)

This example will validate this procedure by comparing the results to those in Lachin (2011) on page 351. In this example, OR = 0.4444, P_E = 0.15, R^2 = 0.0, M_D = 1, M_H = 2, power = 0.85, and alpha = 0.05. The test is two-sided. The resulting sample size is 161.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Odds Ratio in a Matched Case-Control Design with a Binary X** procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size N**
Alternative Hypothesis	**Two-Sided**
Power...	**0.85**
Alpha..	**0.05**
M_D (Number of Cases per Set)	**1**
M_H (Number of Controls per Case).........	**2**
OR (Odds Ratio)	**0.4444**
P_E (Probability of Exposure)	**0.15**
R^2 (Exposure vs. Covariates)	**0.0**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results
Hypothesis Type: Two-Sided

Power	Number of Matched Sets N	Cases per Set M_D	Controls per Set M_H	Odds Ratio OR	Probability of Risk Exposure P_E	Regression of Exposure on other Covariates R^2	Alpha
0.8500	161	1	2	0.4444	0.1500	0.0000	0.0500

PASS has also calculated N to be 161.